

基因芯片技术分析前列腺癌发展相关的基因及功能初探

学校编码: 10384

分类号_____密级_____

学号: 200326069

UDC_____

厦门大学

硕士学位论文

基因芯片技术分析前列腺癌发展相关的基因 及功能初探

Searching Differential Expression Gene in Prostate Cancer
Cell Lines Using Microarray and Analysis of Gene Function

涂智杰

指导教师: 陈亮 教授

周建光 教授

专业名称: 细胞生物学

论文提交日期: 2006 年 4 月 25 日

论文答辩时间: 2006 年 6 月 5 日

学位授予日期:

答辩委员会主席: 陶涛 教授

评阅人: _____

2006 年 6 月

厦门大学学位论文原创性声明

兹呈交的学位论文，是本人在导师指导下独立完成的研究成果。本人在论文写作中参考的其他个人或集体的研究成果，均在文中以明确方式标明。本人依法享有和承担由此论文产生的权利和责任。

声明人（签名）：

2006 年 月 日

厦门大学学位论文著作权使用声明

本人完全了解厦门大学有关保留、使用学位论文的规定。厦门大学有权保留并向国家主管部门或其指定机构送交论文的纸质版和电子版,有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆被查阅,有权将学位论文的内容编入有关数据库进行检索,有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

本学位论文属于

1、保密 (), 在 年解密后适用本授权书。

2、不保密 (√)

(请在以上相应括号内打“√”)

作者签名:

日期: 年 月 日

导师签名:

日期: 年 月 日

目录

中文摘要.....	1
英文摘要.....	4
1、前言.....	7
1.1 前列腺癌背景介绍.....	7
1.2 Affymetrix 芯片介绍.....	8
1.3 运用 Affymetrix 芯片整体研究前列腺癌细胞系表达谱.....	10
2、材料与方法.....	11
2.1 材料.....	11
2.2 方法.....	13
3、结果.....	24
3.1 芯片结果.....	24
3.2 反向狭缝杂交、传统 RT-PCR 和实时荧光 PCR 验证芯片结果.....	31
3.3 候选基因的功能初探.....	40
4、讨论.....	47
4.1 芯片结果分析.....	47
4.2 反向狭缝杂交、传统 RT-PCR 和实时荧光 PCR 验证芯片结果分析.....	48
4.3 候选基因的功能初探分析.....	48
5、小结.....	50
6、参考文献.....	51
7、附录.....	56
8、致谢.....	66

CONTENTS

Chinese abstract.....	1
English abstract.....	4
1. Introduction.....	7
1.1 Background of prostate cancer.....	7
1.2 Affymetrix microarray manual.....	8
1.3 Microarray analysis of differential gene expression in prostate cancer cell lines.....	10
2. Materials and Methods.....	11
2.1 Materials.....	11
2.2 Methods.....	13
3. Results.....	24
3.1 Results of Microarray.....	24
3.2 Using Reverse slot blotting, Traditional RT-PCR and Real-time PCR methods to verify the results of Microarray.....	31
3.3 Candidate gene function study.....	40
4. Discussion.....	47
4.1 Analysis of Microarray results.....	47
4.2 Analysis of the methods of verification.....	48
4.3 Analysis of Candidate gene function study.....	48
5. Summary.....	50
6. Reference.....	51
7. Appendix.....	56
8. Acknowledgement.....	66

摘要

人类基因组测序工作的完成使人们可以方便调用任何基因序列,但仅有基因序列并不能解释众多的生物学问题,这要求人们发展一种高通量的技术用于研究基因的生物学功能以及与其他基因相互作用的关系。DNA 微阵列技术以其高通量的特点已经在肿瘤生物学的研究中逐渐被采用。由于癌症是源于基因表达谱改变的一种基因疾病,通过 DNA 微阵列技术研究代表肿瘤发展各阶段细胞之间的基因表达差异将会使人们更好的了解肿瘤的形成和发展过程。

前列腺癌是引起西方男性死亡的主要癌症,而且也是最常见的诊断性癌症。通常,前列腺癌变的速度很慢。如果早期发现并治疗可以得到很好的疗效,然而一旦前列腺癌出现转移就很难进行有效治疗和提高病人的生存率。为了解决上述问题,关键是使用预测性标记物找出哪些前列腺癌病人在其有生之年会进一步发展为转移型前列腺癌。转移型前列腺癌经过了一系列的发展过程其中包括正常前列腺上皮细胞内瘤、局部前列腺癌、侵袭性前列腺癌和转移性前列腺癌。前列腺癌的这些发展阶段包括多个分子的改变,暗示这些发展是通过基因表达差异的改变进行的。

一旦前列腺癌发展为雄激素非依赖性就意味着这些肿瘤细胞的生长不受控制,这个阶段是前列腺癌病人死亡的主要原因。然而几乎所有的转移型前列腺癌病人起初对抗雄激素治疗都有效果,但是在 2 年内大部分的病人就对抗雄激素治疗失去应答。基因表达的改变和表观遗传的改变被认为是前列腺癌发展为雄激素非依赖性的主要原因^[1-4]。为了研究这种发展过程,我们选择使用微阵列技术来研究高恶性、晚期、雄激素非依赖性的前列腺癌细胞系 C4-2 和低恶性、初期、雄激素依赖性前列腺癌细胞系 LNCaP 基因表达图谱差异。获得了以下研究结果:

1. 选用涵盖 18,000 多个转录本,代表 18,000 多个明晰的基因,其中 13,000 多个为全长基因的美国 Affymetrix 公司的人类全基因组 U133 系列芯片,高通量分析了 LNCaP / C4-2 细胞中基因表达图谱的变化。
2. 通过 Affymetrix、GenBank, dbEST, and RefSeq 等网站对 Affymetrix 芯片数据进行大量的生物信息学分析和文献调研。其中表达量差异在 1 倍以上基因 658 个,表达量差异在 2 倍以上的基因有 260 个。175 个基因在 C4-2 细胞中

上调表达, 85 个基因在 C4-2 细胞中下调表达。其中有些差异表达的基因如: IGF-I、IGFBP3 等基因通过人们的深入研究证明在前列腺癌中起重要作用。

3. 从表达量差异在 2 倍以上的 260 个基因中选取了 118 个基因作为验证芯片结果的候选基因。通过反向狭缝杂交、传统 RT-PCR 和实时荧光 PCR 方法验证候选的 100 个有差异表达的基因发现: 100 个候选基因中有 48 个基因的表达变化趋势跟芯片结果一致。其中在 C4-2 细胞中上调表达的基因有 206183_s_at、214240_at、207245_at、210145_at、209191_at、220085_at、201195_s_at、205476_at、208978_at、212526_at、212192_at、201387_s_at、202388_at、206067_s_at、209505_at、206069_s_at、219630_at、39248_at、202016_at、204693_at、205893_at、209679_s_at、221816_s_at、220625_s_at、205597_at、204971_at、202679_at、206183_s_at、220129_at、201487_at; 在 C4-2 细胞中下调表达的基因有 217564_s_at、206624_at、211682_x_at、206700_s_at、219974_x_at、209348_s_at、213967_at、201909_at、212233_at、203845_at、203675_at、213479_at。
4. 由于前列腺癌细胞的生长和发展与雄激素有关, 因此本实验通过用雄激素 R1881 处理 LNCaP 细胞提取 RNA, 用于研究候选的 48 个基因的表达是否受雄激素的调节。结果表明受雄激素调控的基因有 202016_at、205893_at、209679_s_at、220625_s_at、205597_at、204971_at、201195_s_at、212192_at、202388_at、206069_s_at、217564_s_at、209813_x_at、219974_x_at。
5. 由于工作量的问题, 本实验选取了 205476_at(CCL20)和 217564_s_at(CPS1)两个基因进行功能初探。理由是 205476_at 的表达差异通过实时荧光 PCR 结果它的表达量在 C4-2 中上调表达 51.71 倍, 217564_s_at 是一个核基因编码并定位在线粒体基质中起作用的蛋白并且在 C4-2 细胞中下调表达来了 29.96 倍, 因此把其作为抑癌基因研究其功能。
6. 对于候选基因 205476_at (CCL20) 已经完成真核和原核基因表达载体构建, 并通过测序确认正确。但是把含有 CCL20 的载体稳定转染 LNCaP 细胞发现不能形成稳定克隆, 推测其高表达对细胞有毒性。而且本实验还完成了制备 CCL20 蛋白抗体的质粒 pGEX4T1-CCL20 的构建, 并通过 IPTG 诱导表达证实构建成功。通过掺入 Brdu 检测细胞增殖实验发现 CCL20 的瞬时表达可能对细

胞的增殖有促进作用。

7. 对于 217564_s_at (CPS1) 也完成了真核表达载体构建。由于其下调表达所以把 CPS1 基因稳定转染了 DU-145 细胞, 实验发现虽然 DU-145 细胞可以形成克隆, 并在 DNA 水平上可以检测其已经整合到细胞基因组中, 但是在 RNA 水平上并不能检测出 CPS1 的表达, 推测由于 CPS1 基因编码区全长 4500bp, 可能在基因扩增时引入了突变使基因表达不能正常进行。

虽然芯片结果提供了大量的候选基因, 但是如何分析候选基因的功能成为研究基因功能的主要瓶颈, 对候选基因的功能分析将为人们了解前列腺癌发展的分子机理提供可能。

关键词: 微阵列; 雄激素非依赖型前列腺癌; 实时荧光 PCR

ABSTRACT

The vast amount of information available through the human genome project .However, the mere sequence information of the whole genome does not answer all our questions. What is required at this stage is a complete understanding of the function of genes and other parts of the genome so as to uncover how sets of genes and their products work together in normal and diseased conditions. DNA microarrays are some of the most powerful and versatile tools available, and there are several applications of microarray technology in cancer biology. Since cancer is a genetic disease arising from the progressive accumulation of many genetic alternations, identification of differences in the expression profile of tumor cells in comparison to their normal counterpart would provide a better platform for understanding the process of tumor formation and development.

Prostate cancer is a leading cause of cancer-related deaths and the most commonly diagnosed cancer in men in the west. By nature, cancer in the prostate progresses slowly and can be treated effectively when detected early; however, the metastastatic disease presents a major challenge to improve survival rate and treatment efficacy. To overcome this problem, it is critical to identify predictors to distinguish prostate cancers that will progress and metastasize, and to separate them from those that will not progress during the expected lifetime of the patient. Metastatic prostate cancer proceeds through a series of distinct states such as transformation of normal prostatic epithelial cells to preinvasive primary tumor, androgen-dependent invasive cancer, and androgen-independent (AI) metastatic disease. These stages of prostate cancer involve multiple molecular changes some of which can be implicated to alterations in gene expression.

Progression to androgen independence (AI) leading to uncontrolled cell growth is the main cause of death in prostate cancer. While almost all patients with metastatic prostate cancer will initially respond to anti-androgen treatments, the majority will fail hormonal treatments in less than 2 years. Both genetic and epigenetic alterations in

gene expression contribute significantly to the development of AI.

To investigate this we have used an in vitro cell line model of AI prostate cancer C4-2 and androgen dependent prostate cancer LNCaP from which we have identified a number of differentially expressed genes associated with progression to AI in prostate cancer. Our results show that:

1. Using bioinformation analysis found that 658 genes were differential expressed; 260 genes were differential expressed by at least two fold. 175 genes upregulated and 85 genes downregulated in C4-2. Such as IGF-1、IGFBP3 gene was included in the microarray results. These genes were extensively studied.

2. In order to verify the results of microarray, we use the method of traditional Reverse Transcription PCR; Reverse Slot Blotting; Real time PCR. We had chosen 118 genes on the basis of differential expression fold. In the verification, we found 48 genes are the same as the results of microarray. 206183_s_at、214240_at、207245_at、210145_at、209191_at、220085_at、201195_s_at、205476_at、208978_at、212526_at、212192_at、201387_s_at、202388_at、206067_s_at、209505_at、206069_s_at、219630_at、39248_at、202016_at、204693_at、205893_at、209679_s_at、221816_s_at、220625_s_at、205597_at、204971_at、202679_at、206183_s_at、220129_at、201487_at gene upregulated in C4-2; 217564_s_at、206624_at、211682_x_at、206700_s_at、219974_x_at、209348_s_at、213967_at、201909_at、212233_at、203845_at、203675_at、213479_at gene downregulated in C4-2.

3. 202016_at、205893_at、209679_s_at、220625_s_at、205597_at、204971_at、201195_s_at、212192_at、202388_at、206069_s_at、217564_s_at、209813_x_at、219974_x_at gene expression were influenced by R1881.

4. In the experiment I had chosen 2 genes for their functional analysis, we constructed five plasmids successfully.

5. I found that transfect the CCL20 to the LNCaP could not generate the stable clone. We supposed that the CCL20 excessive expression would poison the cell. We found CCL20 could promote the cell proliferation in the transient transfection under the method of Brdu incorporation.

6. Transfect CPS1 to the DU-145 could generate the stable gene but the gene did not

express, we supposed that CPS1 had mutation.

7. The results of microarray provided a number of differential expressed genes, but these genes functional analysis was the main difficulty. If we could study the differential expressed gene function, we would understand the molecular development of prostate cancer.

Key words: microarray; AI prostate cancer; real-time PCR

1. 前言

1.1 前列腺癌背景介绍

在欧美国家，前列腺癌是检出率最高的肿瘤，其致死人数在所有男性肿瘤中居第二位。随着老龄人口的增加和生活环境的改变，我国前列腺癌发病率近年来呈不断上升趋势。但是由于前列腺癌发病隐匿，大约有 60% 的病人在表现出临床症状时已处于癌症晚期。前列腺癌的病因学不太明确，组织病理学变化较多，与内分泌因子关系复杂，术后易复发等因素，为早期检测与临床治疗造成了很大困难。

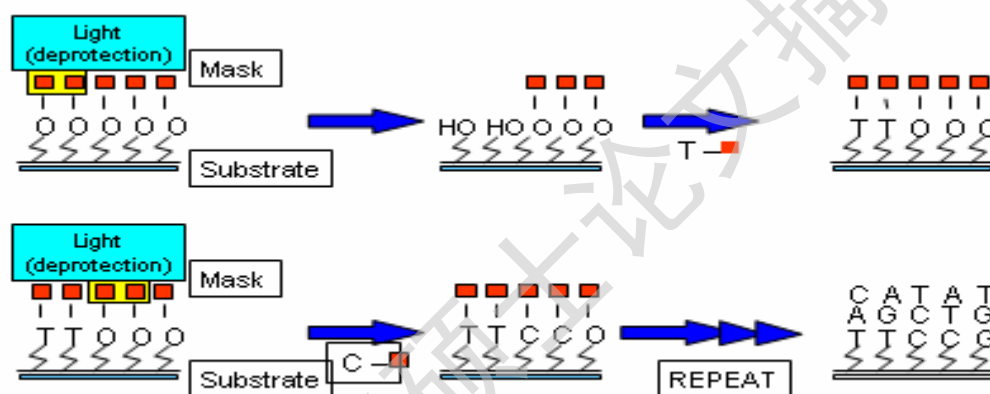
前列腺癌是由前列腺上皮细胞的转化、恶变引起的，这是一个多步骤多阶段的过程，初级阶段肿瘤细胞的生长与存活为雄激素依赖型，撤除雄激素或抗雄激素的内分泌疗法对病人有较好疗效，而恶化程度较高的前列腺癌为雄激素非依赖型，内分泌疗法往往失效。1991 年，美国弗吉尼亚大学钟维国教授实验室建立了模拟人前列腺癌发展的 LNCaP 细胞模型。在该模型中，将成瘤性较低的人前列腺癌细胞系 LNCaP 与骨基质 MS 细胞共同接种裸鼠，八周后，对其进行阉割去势手术，经过一段时间的停顿，LNCaP 细胞和 MS 细胞形成的嵌合肿瘤继续生长，PSA 分泌增高，四周后进行肿瘤摘除，在体外培养，经细胞学和分子生物学进行来源鉴定后得到恶化程度较高的雄激素非依赖的 C4 细胞系，若五周后进行摘除，则命名为 C5 细胞系。将 C4 细胞系再与骨基质成纤维细胞系 MS 共接种至阉割去势的裸鼠皮下，十二周后摘除肿瘤体外培养，则获得雄激素非依赖具转移特征的 C4-2 细胞系切除睾丸获得了雄激素非依赖性的 C4-2 细胞株，该细胞代表了前列腺癌发展、恶化的更高阶段，具有强的致瘤性和转移性。细胞遗传学（荧光原位杂交 FISH、比较基因组杂交 CGH）与分子生物学研究表明，在前列腺癌恶化过程中，细胞的核型及基因表达谱都发生了不可逆的转变。

人前列腺癌细胞模型的建立对前列腺癌细胞的生长、发展及转移的分子生物学研究提供了一套很好的途径，而目前研究认为前列腺癌的发生、发展和恶化可能与前列腺特异表达的基因有关。

1.2 Affymetrix 芯片介绍

(1) 高效准确的光蚀刻原位合成技术

美国著名的 Affymetrix 公司率先开发的寡聚核苷酸原位光刻专利技术，是生产高密度寡核苷酸基因芯片的核心关键技术。其基本原理：首先使固相片基羟基化，并用光敏保护集团将其保护起来，然后选取适当的闭关膜（mask）使需要聚合的部位透光，其它部位不透光。这样，当光通过蔽光膜照射到支持物上时，受光部位的羟基就会发生脱保护而活化，即可结合碱基。由于合成的碱基单体一端是用于固相合成，另一端受光敏集团的保护，所以原位合成后，可进行下一轮的光照、脱保护和固相合成。与此循环下去，不断改变蔽光膜的通透位点，就可以实现在同一玻片上合成成千上万种预定序列的寡核苷酸探针。（图 1）



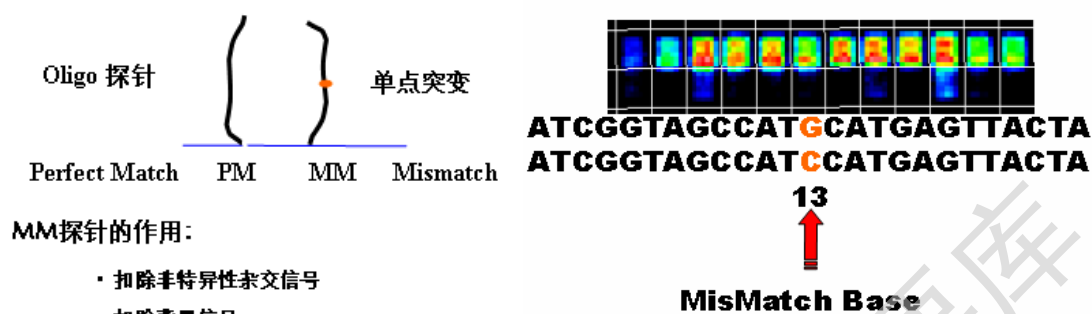
该方法的最大特点在于用很少的步骤可合成大量的 DNA 阵列。如要合成一段含 N 个碱基的寡核苷酸序列只需 $4 \times N$ 个化学步骤，可同时合成 4^N 个可能结构。如果用传统的先合成后点样的方法，其工作量的巨大是不可思议的。Affymetrix 的原位合成技术可制作的点阵密度高达 $10^6/\text{cm}^2$ 。

Affymetrix 又将光引导合成技术和半导体工业所用的光敏抗蚀技术相结合，以酸作为去保护剂，使产率增加到 98%（光合成法产率低，不到 95%），同时解决了由于掩膜透光孔径太小而产生的衍射问题，有效的提高了点阵的密度，如果利用波长更短的物质波，如电子射线进行脱保护，点阵的密度可以高达 $10^{10}/\text{cm}^2$ 。

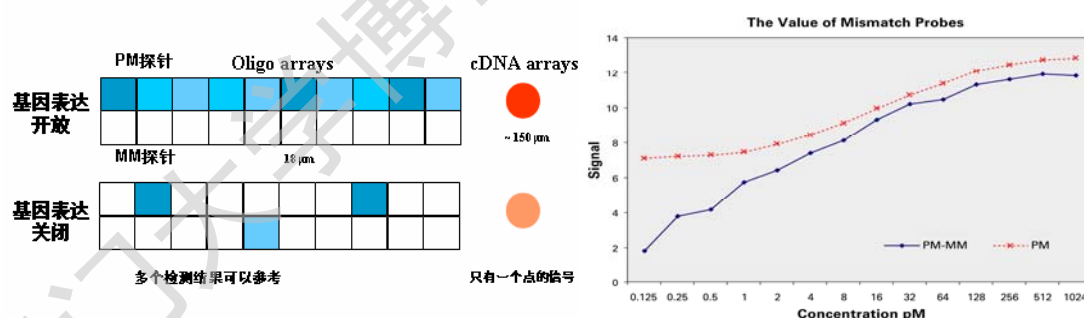
(2) 独特的 PM-MM 探针设计

基因芯片杂交的灵敏度和特异性是芯片技术的核心，Affymetrix 在探讨了各种各样的影响因素后，设计出了一种独特的 PM-MM 探针方案。芯片上的每一个基因或 EST 都是由 16-20 对 25mer 的探针组成，其中一个是完全匹配（perfect

match PM) 的, 另外一个序列中间有一个碱基错配 (mismatch MM)。该设计, 可提高探针的灵敏度和特异性。(图 2)



相比 cDNA 芯片和单一序列的寡核苷酸芯片, Affymetrix 设计多个短的探针片段, 可以有效的区分有同源性的基因序列, 克服了背景噪声、错误和误差, 避免了同源性靶序列与探针交叉杂交, 而引起的假阳性和判断错误。25mer 长度的寡核苷酸的信号强度和分辨率达到了一个平衡, 在这个长度下, 有一个碱基发生错配, 杂交复合物就会变得很不稳定。如果长度达到 60mer, 就无法区别序列上十分接近的序列了。此外, 从多个探针位点检测的荧光信号, 经过综合评估、统计计算和分析, 获得的数据比单个探针判断样品是否存在某一靶序列的数据更为可靠。(图 3)



在 PM-MM 探针设计中, MM 探针是有效的内参照, 它们与 PM 探针一样可以和非特异性序列结合, 这样就可以将不同来源的样品中的背景信号有效的定量扣除。这种独特的设计对于区分特异性和非特异性杂交是相当灵敏的。比较那些单一的基因探针来说, PM-MM 探针的高特异性和灵敏度更适合检测低丰度表达的基因。

Affymetrix 的基因芯片分析平台不仅可以分析某一样品中基因表达的绝对水平, 还可以比较不同样品之间各种基因表达的相对比例, 而双色法需在每次杂交检测中设置对照, 只能进行两个样品之间的比较。Affymetrix 进行多个样品之间

的比较时，只需在多个样本中设置对照即可。

Affymetrix 公司的人类全基因组 U133 系列芯片中的 HU133A 涵盖 18,000 多个转录本，代表了 18,000 多个明晰的基因，其中 13,000 多个为全长基因；序列经过精心筛选，来自于 GenBank, dbEST, and RefSeq；有利于高通量并行分析人类基因表达的变化。

1.3 运用 Affymetrix 芯片整体研究前列腺癌细胞系表达谱

本室周建光等人运用 Affymetrix 公司的人类全基因组 U133 系列芯片中的 HU133A 和上述前列腺癌模型用来研究这两个细胞系在正常培养下的基因表达差异，目的旨在通过基因芯片的高通量的特点，从整体上研究不同细胞在基因水平上的表达差异，从中找出可能起关键作用的新基因，为靶向基因治疗提供更好的选择；从海量的信息中筛选出有重大意义的基因用于研究前列腺癌的发生过程，进一步使人们更好的治疗此项疾病，提高治疗效果及预后生活质量。

通过生物信息学分析 LNCaP 细胞和 C4-2 细胞芯片结果，其中共有 658 个基因有差异表达，并且表达量差异在 4 倍以上的基因有 260 个；其中 175 个基因在 C4-2 细胞中上调表达，85 个基因在 C4-2 细胞中下调表达。由于工作量太大，本人只选用了其中的 2 个基因作为本实验的候选基因，并对其做一些基因功能的初步分析。

2. 材料与方法

2.1 材料

2.1.1 引物合成及序列测定

引物和序列测定由上海博亚生物技术有限公司和上海生工完成，引物序列见附录。

2.1.2 细胞株、菌株

大肠杆菌 E.coli DH5a、BL₂₁，黑色素瘤细胞 B16、结肠癌 Caco-2、鼠成纤维细胞 NIH3T3、肝癌细胞 HePG2、乳腺癌细胞 MCF-7、人前列腺癌雄激素依赖型细胞系 LNCaP、人前列腺癌雄激素非依赖型细胞系 C4-2、宫颈癌细胞 Hela、人前列腺癌脑转移细胞系 DU-145、人前列腺癌骨转移细胞系 PC-3、肺癌细胞 PLA-801。上述细菌和细胞都为实验室所有。

2.1.3 质粒

真核表达质粒 pcDNA3.1 和 pIERS2-EGFP、原核表达质粒 pGEX-4T1 均为实验室保存。

2.1.4 分子生物学工具酶

限制性内切酶	Biolab 公司、Takara 公司
普通 Taq DNA 聚合酶	上海申能博彩生物公司
T4 DNA 连接酶	Takara 公司
Pfu DNA 聚合酶	上海申能博彩生物公司
LA Taq DNA 聚合酶	Takara 公司
Super Taq DNA 聚合酶	上海申能博彩生物公司

2.1.5 常用分子生物学试剂盒

质粒提取试剂盒（小量提取）	Promega 公司
DNA 胶回收试剂盒	Promega 公司
PCR 产物回收试剂盒	Promega 公司

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库